

Semantic Search Engine Based on Query Expansion with Google Ranking and Similarity Measures

Authors : Ahmad Shahin, Fadi Chakik, Walid Moudani

Abstract : Our study is about elaborating a potential solution for a search engine that involves semantic technology to retrieve information and display it significantly. Semantic search engines are not used widely over the web as the majorities are still in Beta stage or under construction. Many problems face the current applications in semantic search, the major problem is to analyze and calculate the meaning of query in order to retrieve relevant information. Another problem is the ontology based index and its updates. Ranking results according to concept meaning and its relation with query is another challenge. In this paper, we are offering a light meta-engine (QESM) which uses Google search, and therefore Google's index, with some adaptations to its returned results by adding multi-query expansion. The mission was to find a reliable ranking algorithm that involves semantics and uses concepts and meanings to rank results. At the beginning, the engine finds synonyms of each query term entered by the user based on a lexical database. Then, query expansion is applied to generate different semantically analogous sentences. These are generated randomly by combining the found synonyms and the original query terms. Our model suggests the use of semantic similarity measures between two sentences. Practically, we used this method to calculate semantic similarity between each query and the description of each page's content generated by Google. The generated sentences are sent to Google engine one by one, and ranked again all together with the adapted ranking method (QESM). Finally, our system will place Google pages with higher similarities on the top of the results. We have conducted experimentations with 6 different queries. We have observed that most ranked results with QESM were altered with Google's original generated pages. With our experimented queries, QESM generates frequently better accuracy than Google. In some worst cases, it behaves like Google.

Keywords : semantic search engine, Google indexing, query expansion, similarity measures

Conference Title : ICCIIS 2014 : International Conference on Computational Intelligence and Intelligent Systems

Conference Location : Barcelona, Spain

Conference Dates : August 18-19, 2014